5-14-2021

What Is Wrong with Statistical Discrimination?

Ziyue Sun
Department of Philosophy, Georgia State University

What Is Wrong with Statistical Discrimination?

by

Ziyue Sun

Under the Direction of Andrew Altman, PhD and Andrew I. Cohen, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Statistical discrimination is a form of discrimination that uses statistical inferences about the groups to which individuals belong as grounds for treating them differently. It remains unclear what, if anything, makes statistical discrimination wrong. My thesis argues that statistical discrimination is wrong because, and insofar as, it contributes to existing social injustice. After an introduction to the issues in section 1, section 2 clarifies the concept of statistical discrimination and its differences with non-statistical discrimination. Section 3 discusses different accounts that seek to explain when and why statistical discrimination is wrong. I examine two approaches, one of which regards the wrong of statistical discrimination as part of discrimination in general, while the other conceives of the wrong as distinctive to statistical discrimination itself. I argue the former approach is better. Among different accounts within the approach, the context-based consequentialist account that explains the wrong in relation to existing social injustice is most promising. Section 4 uses racial profiling as an example to illustrate how the account can help explain this hard case of statistical discrimination, calling upon an argument developed by Benjamin Eidelson.

INDEX WORDS: Discrimination, Statistical generalizations, Social injustice, Consequentialism, Racial profiling, Racism

What Is Wrong with Statistical Discrimination?

by

Ziyue Sun

Committee Co-Chair: Andrew Altman

Andrew I. Cohen

Committee: Suzie Love

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2021

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Drs. Andrew Altman and Andrew Cohen, for providing invaluable guidance, advice and support for me during the process of my writing. I am especially grateful to Dr. Altman, whose research on discrimination and racial justice have inspired me to work on this project. Also, I would like to thank Drs. Sandra Dwyer and S.M. Love who gave constructive comments on my draft and pushed me to clarify and strengthen my argument. Finally, I would like to express my gratitude to my mother and my friends for supporting me throughout my difficult times.

TABLE OF CONTENTS

1 INTRODUCTION

Nowadays people, organizations, and governments increasingly rely on statistics and algorithms to make important decisions and policies. While utilizing statistical information might help avoid some forms of subjective bias, it also makes individuals more vulnerable to statistical discrimination, a form of discrimination that uses statistical inferences about the social groups to which persons belong as grounds for treating them differently. Although many people hold that statistical discrimination is objectionable, it is not clear what makes it wrong.

In this thesis, I argue that a context-based consequentialist account is a promising explanation for the wrong of statistical discrimination. The account conceptualizes the wrong of statistical discrimination in terms of its contribution to existing social injustice against disadvantaged social groups. In Section 2, I clarify what statistical discrimination is and compare it with non-statistical discrimination. Section 3 discusses two approaches to account for when and why statistical discrimination is wrong. One approach argues that statistical discrimination constitutes a distinct form of wrong, one not shared by other forms of discrimination, namely, the wrong of imposing a cost on persons for features they do not have. A second approach explains the wrong of statistical discrimination as an instance of a general wrong shared by all forms of discrimination. I argue that the general-wrong approach is preferable and that, among different accounts within this approach, a consequentialist account has notable advantages. In Section 4, I illustrate how the consequentialist account can offer an illuminating way to understand the moral troubling features of the hard cases of statistical discrimination by defending Eidelson's argument against racial profiling.

2 WHAT IS STATISTICAL DISCRIMINATION?

2.1 The Definition for Statistical Discrimination

The term ‘statistical discrimination’ was first used by the economist Edmund Phelps (1972). It initially referred to the behaviors of employers when they made hiring decisions using statistical information about a group to which applicants belong in order to infer what their job productivity would likely be (Rodgers, 2009, p.223). Lippert-Rasmussen (2007; 2011; 2014) generalizes the Arrow-Phelps account, presenting a definition of statistical discrimination that I adopt for this thesis. On his definition, “A policy, P, constitutes statistical discrimination against a certain socially salient group of people, X-people, in relation to non-X-people . . . if, and only if, (i) there is statistical evidence which suggests that X-people differ from non-X-people in dimension, D, (ii) P involves treating X-people worse than non-X-people, and (iii) P is in place because of (i)” (Lippert-Rasmussen, 2007, p.387). Note that this definition is not moralized, in the sense that it leaves open whether statistical discrimination is necessarily wrong.

Accordingly, there are three conditions for actions or policies to be statistical discrimination. First, there is a condition of disadvantageous treatment. The policy or action treats members of a group in a way that disadvantages them in comparison with nonmembers. Disadvantages include both deprivations of access to resources, opportunities and services, and the imposition of costs. The second condition is about socially salient group membership. The group targeted for disadvantageous treatment needs to be a socially salient group. A group is socially salient if “it is important to the structure of social interactions across a wide range of

social contexts" (Lippert-Rasmussen 2007, p.386), such as groups based on sex, race, color, religion, and disability. By contrast, people whose blood type is AB are not a socially salient group. This restriction is necessary because without it any differential treatment imposing disadvantages on some persons, such as colleges accepting some applicants and rejecting others, would be discriminatory.[1] The third condition is about the role of statistical evidence. The treatment must be motivated and grounded (at least partly) by the agent's belief in a statistical correlation between membership in a socially salient group and some other attribute, such as productivity or criminal behavior. Note that there is no criterion in the definition for the statistical evidence grounding discriminatory treatment to be epistemically dubious or biased. Instead, the statistical evidence may accurately describe the relevant difference between social groups and be used in a non-biased way.[2]

2.2 The Differences between Statistical Discrimination and Other Forms of Discrimination

Before moving on to a discussion on what makes statistical discrimination wrong, if and when it is wrong, it is worth taking a look at the difference between statistical and other forms of discrimination. Suppose there are two cases of discrimination (C1) and (C2).

Case (C1): The employer of a company rejects a female applicant because he receives and believes the statistical evidence from company research showing that the average productivity (measured by output) of female employees is lower than that of male employees.

[1] There is a debate about which socially salient groups count in discrimination and whether there should be a group restriction at all (see Altman (2020) and Thomsen (2013)), but I bracket such complexity in this paper.
[2] By "non-biased" I mean all available relevant statistical evidence is taken into account and the information is not used in a selective way.

Case (C2): The employer of a company rejects a female applicant because the sacred text of his religion says that God does not want women to work outside of home.

Both employers in (C1) and (C2) engage in discrimination against women. The employer in (C1) engages in statistical discrimination according to our definition, but the employer in (C2) does not. The main difference between statistical discrimination and non-statistical discrimination is in the third condition spelled out at the end of 2.1, namely, the role of evidence. While statistical discrimination is motivated and grounded by belief in statistical evidence about the correlation between social group membership and particular attributes, non-statistical discrimination is motivated by other attitudes, such as hostility towards certain social groups or belief in the inferior moral status of the groups. But what difference does the use of statistical evidence make to the nature of statistical discrimination?

First, statistical discrimination is arguably less likely than other forms of discrimination to involve an intention to disadvantage the members of the group in question.[3] It is true that, as Pauline T. Kim points out, statistical discrimination can be intentional, as when the agent relies on algorithms to make decisions because "it knows the model produces a discriminatory result and intends that results to occur" (Kim, 2016, p.884). However, usually, the agent engaging in statistical discrimination does not intend to disadvantage the group. Given the same statistical estimation of productivity, the employer in (C1) might be equally willing to hire female as male applicants. In fact, one reason given for using statistics in employment is to reduce intentional discrimination in decision-making.

Second, the use of statistical evidence introduces a form of uncertainty that non-statistical discrimination is less likely to have. The statistical evidence is almost always probabilistic,

[3] Not all form of non-statistical discrimination involves discriminatory intention, but many of them do.

where the probability is less than 100% (otherwise the grounds for the treatment would not be statistical). Therefore, it is almost always possible that a member of the group does not share the attributes as predicted, and the discriminator will typically acknowledge this empirical possibility, because he knows that he is relying on statistical evidence. If the employer in (C1) rejects all female applicants based on the evidence that 80% of female employees are less productive than an average male employee, he knows that he might dismiss a female applicant who belongs to the remaining 20%, and thus, from his own perspective, he would recognize that there is a non-negligible probability that he has made a mistake and failed in his attempt to maximize productivity. By contrast, the decision of the employer in (C2) does not rest on probabilistic inferences of from statistical premises: his sacred text says that no women are allowed by God to work outside of home and that is the end of the matter for him, he will not acknowledge that he has made a mistake in dismissing a female applicant.

3 WHAT IS WRONG WITH STATISTICAL DISCRIMINATION?

In this section, I compare some influential accounts of what makes statistical discrimination wrong, when it is wrong. Those accounts can be divided into two approaches. The first argues that there is something distinctive about statistical discrimination that constitutes a different kind of wrong from discrimination in general. I call it the "distinct-wrong" approach. The second approach contends that statistical discrimination is the same kind of wrong as discrimination in general, which I term as the "general-wrong" approach.

3.1 A Minimal Standard for A Satisfactory Answer

Before evaluating which approach does a better job, a question to be considered is: What counts as a good account of the moral wrong of statistical discrimination? The problem arises

because supporters of different normative ethical theories, such as deontology and consequentialism, may disagree about what counts as a good account. Debate about what makes statistical discrimination wrong can partly stem from the debate on normative ethical theories, and therefore may seem unresolvable without judging which normative ethical theory is better. It makes no sense to argue for the superiority of a consequentialist account of what makes statistical discrimination wrong, over other explanations, just based on consequentialist ethical principles. Such arguments cannot convince those who commit to different ethical theories. Given that this thesis cannot and does not intend to settle the long-standing debate between different normative ethical theories, a set of evaluation criteria independent of any specific normative ethical theory is needed for the project of the thesis. At the same time, I assume that the different theories that are part of the philosophical tradition all have some plausibility, even if the question of which is best remains unsettled.

To answer the question about what makes statistical discrimination wrong is to find a property P such that P can explain why instances of statistical discrimination are wrong when they are wrong. I propose the following three criteria for candidate property P.

(S1) P is present in paradigm cases of wrongful statistical discrimination, where "paradigm cases" refer to cases that most people would agree to regard as statistical discrimination that is morally wrong.

(S2) P is absent in cases that can be uncontroversially regarded as morally unproblematic.[4]

(S3) P is plausibly regarded as a wrong-making property.

[4] Note that it does not imply all cases where P is absent are not morally wrong, because there may be other properties that can make a treatment wrong.

The three criteria are in no way a complete set of criteria for an ideal account of the moral wrong of statistical discrimination, but they can serve as a minimal threshold. Meeting those criteria does not make an account satisfactory[5], but the failure to meet them suffices to prove an account as unsatisfactory. If one account can meet the criteria, whereas another account fails to do so, then we have reasons to prefer the former account to the latter one.

One problem for applying the criteria is to determine what counts as a paradigm case of wrongful statistical discrimination. Some instances of statistical discrimination are controversial in terms of whether they are morally wrong or not, and as a result they cannot serve as a paradigm test case. Recall the case (C1) I discussed in last section.

(C1) Other things equal, the employer of a company rejects a female applicant because he receives and believes the statistical evidence from company research showing that the average productivity (measured by output) of female employees is lower than that of male employees. (Let's assume that the main cause for the difference is that females undertake more housework and childcare after going home, which results in a decrease in their energy at work, and that the employer knows that this is the cause.)

I suppose most reasonable people will agree that the behavior of the employer in the case is morally problematic. It is possible that some may perceive the treatment as legitimate, but I believe those will be in the minority and it is difficult to offer a forceful justification for the treatment. Therefore, I will use (C1) as a paradigm case of wrongful statistical for my analysis. In the next section, I will apply the above set of criteria to assess different approaches to the question of what makes statistical discrimination wrong.

[5] There are different ways for an account to be "satisfactory". I regard an account as satisfactory if it can give an explanation for when and why statistical discrimination is wrong that is consistent, can capture the moral significance of the phenomenon, and does not deviate too far from people's moral intuitions.

3.2 The Distinct-Wrong Approach

The first approach to explain the wrong of statistical discrimination is the distinct wrong approach. This approach claims that the wrong of statistical discrimination is different or partly different from the wrong of non-statistical discrimination.[6] In other words, there is something distinctive about statistical discrimination that makes it wrong, regardless of whether and why discrimination in general is wrong. I have shown in my previous discussion that any difference between statistical discrimination and other forms stems from the difference in the role of statistical evidence. If the distinct-wrong approach is right, then there is something wrong with using statistical evidence to ground differential treatment.

One argument adopting this approach is that statistical discrimination imposes costs on one individual for the features of others. The principle underlying this argument is that one can only legitimately be held accountable for all and only those events for which she is responsible. A person is usually not responsible for the features of others. Consider the argument of Colyvan et al. (2001), who point out that convicting someone based on evidence about the behavior of others who are believed to be similar to the person in relevant respects, instead of evidence about the person's own previous behavior, is unfair and unjust. Statistical discrimination seems to commit a similar wrong. To see how this is the case, consider the paradigm case (C1). Suppose the statistical evidence the employer receives shows that 80% of female employees are less productive than an average male employee and the statistical evidence is accurate. Still, 20%

[6] Whether the wrong of statistical discrimination is entirely or partly different from the wrong of non-statistical discrimination depends on which theory about the wrong of non-statistical discrimination to take. Some might think that the wrong of non-statistical discrimination consists in its disadvantageous consequences, which are shared by statistical discrimination. Then there are two parts in the wrong of statistical discrimination, the part shared by all forms of discrimination and the part distinct to statistical discrimination, for which it is *partly* different from the wrong of general discrimination. Others might think that the wrong of non-statistical discrimination consists in discriminatory intention or other attributes that statistical discrimination does not necessarily share. Then the wrong of statistical discrimination can be *totally* different from the wrong of general discrimination. Still, both views admit that part of the wrong of statistical discrimination is distinct to itself.

female applicants are in fact at least as productive as an average male employee. Then the female applicants who are rejected for the job but in fact belong to the 20% are disadvantaged for the low productivity of other women rather than for their own productivity. This seems to be a wrong to those women.[7]

One might ask why this is a wrong that is distinctive to statistical discrimination, different from the wrong of general discrimination. This is because discrimination that is not grounded on statistical evidence does not necessarily impose costs on persons for features they do not have. Unlike the employer in (C1), the employer in (C2) does not disadvantage some female applicants for the features of other women; rather, he dismisses female applicants because of his belief that he should follow what his sacred text says about the role God assigned to every woman.

It is important to note that, in fact, the wrong of punishing or otherwise disadvantaging persons for the features of others does not rest on *differential* treatment of social groups. *Equal* treatment of different social groups can commit the same wrong. For instance, suppose an employer reduces the salary of all employees based on the evidence that half of the employees leave early. In this case, the treatment is equal: all employees, both male and female, suffer undue cost. Still, half of the employees who do not leave early would be punished for the behavior of others, which is also unfair. Thus, it seems that, according to Colyvan's distinct-wrong account, what makes statistical discrimination wrong is its reliance on statistical generalization rather than its discriminatory feature, that is, its feature of unjustifiably treating some worse than others.

[7] An account distinct from Colyvan's account would claim that the reliance on statistical inferences can be morally nonproblematic when it relies on all available information that can be acquired without unreasonable cost. Such account is superior to Colyvan's account and could play a role in a mixed consequential-deontological theory, however I do not have space to further explore it.

Additionally, the explanation does not distinguish wrongful statistical discrimination from morally acceptable uses of statistical information. Specifically, there are cases that rely on statistical inference and are likely to impose costs on persons for features the person does not have, but those cases are normally regarded as morally acceptable. This is incompatible with the criterion (S2) for an acceptable account of the wrong of discrimination.

To see how it is the case, consider other generalizations we use in our lives. All forms of generalization that infer the attributes of an individual from the attributes of other similar individuals seem to involve the possibility of error. Any decision that relies on generalization is likely to disadvantage some persons for attributes that they are predicted to have but actually do not have. We routinely rely on statistical generalization. When we make a decision regarding persons, there is always a property or a set of properties (property X) that is essential for making the decision. However, often we cannot observe directly whether the persons in question have property X. So we need to rely on some proxy property Y that is correlated with property X to make an inference about property X. Whenever we do so, we are relying on statistical generalization. But we do not think reliance on statistical generalization is necessarily unjustifiable. For instance, most universities require applicants to have a high school diploma or equivalent based on the belief that there is strong correlation between high school diplomas and a good command of the knowledge and skills that are necessary for studying in universities. It is possible that there are some applicants without a high school diploma who can do as well as high school graduates. In this sense, such an admission policy imposes cost on those applicants, which might seem unfair. Still, few people find the universities' policy unreasonable. Since this account based on disadvantaging persons for the features of others does not distinguish wrongful

instances of statistical discrimination from morally acceptable ones, it fails to satisfy criterion (S2) and so should be judged inadequate.

3.3 The General-Wrong Approach

Compared with the distinct-wrong account, a better approach might be to explain the wrong of statistical discrimination as part of general discrimination. According to this approach, what makes statistical discrimination wrong, when it is wrong, is not the role of statistical evidence, but something shared with other forms of wrongful discrimination. If so, it seems that the question about what is wrong with statistical discrimination can be reduced to the question about what is wrong with discrimination in general. Then all we would need to do is to apply the theory about the wrong of discrimination per se to statistical discrimination. However, as my following analysis will show, not all theories about the wrong of discrimination in general can give a satisfactory account of the wrong of statistical discrimination. This is because while the use of statistical evidence does not make statistical discrimination intrinsically wrong, it may still make a moral difference, depending on how the wrongness of discrimination is understood. In what follows, I examine some accounts of the wrong of discrimination in general and argue that the accounts based on contribution to existing injustice have notable strengths compared with other accounts.

3.3.1 The Account Based on Merit

Some scholars argue that discrimination is wrong because it violates the norm that we must treat individuals based on their merits. An advocate for this account is David Miller, who maintains, “Justice demands that the job be offered to the best-qualified applicant. We express this by saying that the best-qualified applicant deserves the job or, in a slightly different formulation, that the principle involved is one of merit. This is the principle that condemns

discrimination on grounds of sex, race or religion when hiring employees" (2001, p.156). However, Miller's account is flawed. It is not clear what the criteria for judging who is the best-qualified are, and there might be different reasonable criteria that contradict each other. For example, there might be different criteria for the job of high school teacher, depending on different understanding of the role. Some may believe that the best-qualified teacher should be flexible in grading and reward not only students' performance but efforts, while others may believe that the best-qualified teacher should take only students' performance into account when grading to avoid bias. It is debatable which criterion is better, but both seem reasonable. Besides, this account does not apply to discrimination in areas that are not supposed to be competitive, such as discrimination in policing. That is to say, violation of merit is unable to explain all paradigm cases of discrimination.

Moreover, this account is especially problematic when it comes to statistical discrimination. Statistical discrimination in employment often results from the attempt to find out who is best qualified for a job. To illustrate, in the paradigm case (C1), the employer who rejects women because of the statistical correlation between gender and productivity seems to be pursuing rather than violating the norm requiring the choice of the best qualified (in this case, the most productive) applicants, but his act still seems morally problematic. It suggests that obeying the norm of merit does not preclude some instances of statistical discrimination from being wrong, because violation of the norm of merit is absent in some paradigm cases of wrongful statistical discrimination. Accordingly, the account based on merit fails to meet criterion (S1).

Some might say the employer is not really obeying the norm of merit. The point is more obvious if the number is changed downward. An employer who skips looking at all women candidates based on evidence that 10% percent of women are less qualified than men, can hardly

be regarded as following the norm of merit, although he can claim he is hiring for merit. My reply will be that numbers matters for the norm of merit. To obey the norm of merit, it is required that the treatment is based on *cogent* generalization about merit. The generalization in the revised example is not cogent, which arrives at the conclusion that women are likely to be less qualified than men based on the premise that 10% of women are less qualified than men. However, pursuit for merit can be grounded on generalization that is not, at least obviously, non-cogent, thus accords to the norm of merit but is still morally problematic.

3.3.2 The Account Based on Objectionable Mental States

Another account of the wrong of discrimination is the one based on objectionable mental states. The underlying principle for this account is that "what makes an action morally right depends on whether it expresses the appropriate valuations of…persons," where "expresses" means "manifests a state of mind" (Anderson & Pildes, 2000, pp.1504-1506). Accordingly, what makes discrimination wrong is the objectionable attitudes it expresses. Different versions of this account ground the wrong of discrimination on different mental states, such as "unwarranted animus and prejudice" (Arneson, 2006, p.779) or "unwarranted contempt" (Cavanagh, 2002, p.166), and the belief that some social groups are morally less worthy than others (Alexander, 1992; Glasgow, 2009; Hellman, 2018).

This type of view also falls short in accounting for the wrong of statistical discrimination, because wrongful statistical discrimination does not necessarily involve objectionable mental states. That is to say, criterion (S1) is unmet. Does the employer in (C1) have any contempt or prejudice against women, or hold that they are morally inferior to men? It is possible that he does, but not necessarily. There is nothing contradictory in believing that women are of equal moral worth to men and also believing the truth of a piece of statistical evidence from a reliable

source that women generally have low productivity for certain jobs. It is even logically consistent to believe in the correlation between social group membership and unfavorable traits at the same time as believing that the social group in question has a *high* moral status. For instance, male chauvinists who believe that men have more moral worth than women can consistently believe that men are less likely to succeed in nursing.

A reply to my criticism might be that, although taking belief in the statistical correlation between a certain social group and unfavorable traits does not presuppose prejudice or belief in unequal moral status, acting on such a belief reflects *indifference* to the interests of members of the disadvantaged group, which is also an objectionable attitude. For example, if an employer who believes the evidence that women are less likely to succeed in a job requiring great physical strength cares about the interests of women, then he should find ways to reduce the amount of physical strength required by the job (e.g., implementing new female-friendly machines), rather than reject female applicants. I think this reply is not very convincing. First, not taking measures to promote the interests of women does not necessarily express indifference to women's interests. The employer may wish he could do something rather than rejecting female applicants, but it is just too expensive to do so. Besides, although it is morally praiseworthy if an employer takes positive measures to reduce the obstacle for women to succeed in a job at his own cost and it may sometimes be obligatory for the employers to do so, undertaking such measure is not generally a moral requirement on employers. In many cases those measures would involve undue costs to the employers.

Overall, the account based on objectionable mental states cannot adequately explain the wrong of statistical discrimination.

3.3.3 Discrimination as Contributing to Existing Injustice

Instead of focusing on the subjective attitudes underlying discriminatory behaviors, an alternative way to account for the wrong of discrimination is by looking at its consequences.

It is useful to compare discrimination and differential treatment that is not commonly considered discriminatory. Consider the following two scenarios: (1) An employer rejects all applicants with blood type AB because he personally dislikes persons with blood type AB, and (2) An employer rejects all female applicants because he personally dislikes females. The behavior of the employer in (1) does not count as discrimination under our definition, because "persons with blood type AB" is not a socially salient group. Both scenarios involve wrongful behavior, but it seems that the behavior of the employer in (2) is wrong in a distinct way, because it is more pernicious. Given that the underlying mental states are similar (i.e., dislike for a particular group), it is not the mental states per se that explain the difference. Then what does explain the difference? I think it is the outcome of the treatments. The treatment in (2) disadvantages the victim group in a distinct way, both economically and emotionally. Whereas the applicants with type AB blood are rejected in (1) are likely to find a job elsewhere, the female applicants may find themselves facing many other obstacles in job markets. The rejected applicants in (1) might feel disappointed, embarrassed and annoyed at the employer, but it would not be reasonable for them to experience strong feelings of stigma and powerlessness for being a member of the group as the female applicants in (2) do.

The comparison suggests that the wrong of discrimination is determined by social context, prominently including the historical injustice suffered by particular social groups. While women have suffered systematic and structured oppression throughout history, persons with blood type AB have not. Differential treatments targeted at those groups have the effect of

reinforcing existing injustice towards the victim groups, which makes discrimination morally wrong. Thomas Scanlon holds a view along this line of thought. He points out that the distinctiveness of discrimination is that "the prejudicial judgments it involves are not just the idiosyncratic attitudes of a particular agent but are widely shared in the society in question and commonly expressed and acted on in ways that have serious consequences (Scanlon, 2009, p.73)." That is to say, discrimination adds to the widespread biases towards historically subordinated group, and actions based on such biases perpetuate such injustices and risk worsening the position of persons already suffering undue disadvantage.

In my view, this theory about the wrong of general discrimination provides a plausible account of what makes statistical discrimination wrong, if we consider how statistical discrimination takes place and how it does harm to the society.

Many of the statistical correlations that motivate and are used to ground statistical discrimination come into existence due to injustice in the past. For instance, the apparent correlation between black people and high rates of crime[8] on which racial profiling is based is not a product of nature, but "as a result of the deprivation resulting from discrimination and unjust, racial inequality" (Lippert-Rasmussen, 2006, p.194). Similarly, the supposed correlation between females and low productivity, if it really exists, might be greatly influenced by the "motherhood penalty" (Theunissen et al., 2011). Women of child-bearing age are likely to take long parental leaves, and the disproportionate burden of parental duty has a negative impact on female employee's performance. At first glance, this might seem unrelated to social injustice, since pregnancy is a natural trait of women. However, the natural traits of women are not themselves necessarily disadvantages in employment. They can only constitute disadvantages

[8] The correlation has been pointed out as dubious, because the police crime data may be distorted in itself. Here I set aside such considerations, but it does not mean that I commit to the truthfulness of the claimed correlation.

under certain social conditions. In our society, the relevant social conditions include (1) workplaces routinely fail to provide reliable and affordable childcare, (2) mothers are expected to undertake more responsibility of taking care of children, and (3) paternity leave is not as prevalent as maternity leave in most contraries. As a result, the natural traits of women become employment disadvantages. This is what MacKinnon (1987, pp.32-47) criticizes as a "male standard", where men's physiology and need define the standard way to live and work in our society. Conceive a possible society that is the same as the actual one except that there exists no historical social injustice. There would still be differences between different social groups, but the differences are not likely be so abundant or systematic as to ground statistical discrimination that disadvantages a social group in a systematic way.

Consider one form of statistical discrimination that is increasingly having a great impact on society, discrimination driven by algorithms, i.e., a sequence of well-defined, computer-implementable instructions. Given certain data input, an algorithm can solve problems and make decisions. The use of algorithms is transforming the workplace. Many employers rely on algorithms to determine who gets interviewed and whom to hire or promote (Kim, 2016). Algorithms learn based on the data on which they are designed. Unable to distinguish correlations that represent causal connections from correlations that are not themselves causal but can be explained by a third-party cause or other underlying causal process, algorithms treat instances of prejudice in training data as valid examples and simply perpetuate these injustices in its outputs (Hayes et al., 2020, p.12). For example, suppose an algorithm is trained on a data pool including gender and occupational information of a county's population over the past ten years, and it is asked to predict whether a candidate is likely to succeed in a leadership position. The algorithm can easily detect that being male is correlated with higher possibility of becoming

leaders and make decisions favoring male candidates over female candidates. The algorithm is not able to distinguish the situation where females are intrinsically unsuitable for being a leader and therefore have lower rates of taking leadership position, from the situation where there are third-party factors that unjustifiably make it more difficult for women to get a leadership position. The nature of algorithms is such that decisions based on them reflect and reinforce existing social patterns. Therefore, without reference to such patterns and past injustice, we would not be able to account for wrongful discrimination driven by algorithms.

Statistical discrimination adds to the existing injustice suffered by the subject group. This is realized through a dual mechanism. First, statistical discrimination contributes to existing injustice directly through causing disadvantageous outcomes for the group. For example, statistical discrimination in the workplace increases the difficulty for members of minority groups to get employment, salary raises, and promotion, preventing them from moving out of poverty and improving their living conditions. These consequences, in turn, reduce their access to good education and skill training, which then worsens their performance in the workplace, creating a vicious circle. Statistical discrimination in education, housing, and services also exacerbates the quality of life of minority groups. Besides the material disadvantages, statistical discrimination causes reasonable feelings of powerless, stigma, resentment, hatred and other traumatic experiences.

Second, statistical discrimination also contributes to existing injustice by expressing and affirming the correlation between social group membership and unfavorable attributes, thereby perpetuating certain forms of social injustice in a society with a history of injustice. There are widespread stereotypes and biases against minority social groups. The social meaning of

discrimination in such a society expresses the stereotypes and bias, thereby producing harmful effects on people's attitudes and behaviors towards the target social group.

One might ask how biased expressions produce a harmful influence on society, given that the messages being expressed are something so widespread that almost everyone knows them. If you express in public a stereotype that many people know, does it cause anything that cannot be produced without you expressing the statement? Yes. Your expression works as a confirmation that you endorse this statement, and it also causes other people to know that you endorse this statement. It can have at least two consequences. First, your expression may make the statement sound more convincing to those who do not believe in it and may add to their tendency to start endorsing the statement. Second, even if other people may already endorse this statement before you express it, your expression can strengthen their belief in the statement and make it harder for them to give up their belief in the statement in the future. Biased expression can produce effects that would not be present without it, and this kind of effect can be greater if the message is expressed by a subject with authority or expertise, such as government or other public institutions.

In the case of wrongful discrimination, the messages being expressed reflect unjustifiable biases against certain social group. Those expressions serve to strengthen people's beliefs in such biases and motivate people to act on those beliefs. For example, employers in a society where employment discrimination against women is prevalent tend to offer better salaries, promotions and opportunities to male employees, because such discrimination expresses the message that women are unproductive and incompetent and adds to the employers' beliefs in the message. Sometimes those beliefs can even be unconscious, such as in the phenomenon of implicit bias (Greenwald &Banaji, 1995). An implicit bias is an unconscious attitude where people attribute

certain qualities to all members of a particular group. Because they operate on an unconscious level, such biases are much less controllable than explicit biases. In this way, statistical discrimination based on implicit biases helps existing prejudice and bias proliferate in society, making it even more difficult to eliminate social injustice.

Overall, instead of alternative accounts, we have reasons to prefer a consequentialist account of the wrong of statistical discrimination that conceptualizes it as an instance of the wrong of general discrimination, namely, the wrong of reinforcing existing injustice toward disadvantaged social groups in a given social context. This consequentialist account avoids the pitfalls of other accounts, including those that focus on disadvantaging persons for features they do not have, those that focus on merits, and those that appeal to objectionable mental states. Those accounts have no easy way to explain the wrong of the paradigm case of wrongful statistical discrimination and allow ordinary use of statistical generalization to be justifiable at the same time. The consequentialist account, on the other hand, better tracks the cause and influence of statistical discrimination.

4 RACIAL PROFILING: A HARD CASE

In this section I am going to elaborate on the account of the wrong of statistical discrimination I defended in last section, by examining a specific and controversial form of such discrimination, namely, racial profiling by police. The focus of my examination is a prominent account of profiling given by Eidelson. His account fits well with my analysis of statistical discrimination, because it is also an argument that focuses on the effects of policies on existing racial injustices. I aim to explain and defend Eidelson's argument against racial profiling, as an illustration of how a context-based consequentialist approach can shed light on this hard case of

statistical discrimination. The analysis can serve as further support to the context-based consequentialist account of the wrong of statistical discrimination, by showing how the account reveals and highlights morally troubling dimensions of racial profiling that other accounts may fail to discuss.

I structure this section into four parts. I start by explaining the concept of racial profiling. Next, I survey some non-consequentialist arguments against racial profiling. The third part explains Eidelson's main argument against profiling, and in the last part, I defend Eidelson's account of profiling by responding to some objections to his argument that he did not anticipate or address.

4.1 The Definition of Racial Profiling

A typical definition for racial profiling is proposed by Risse and Zackhauser (2004, p.136), according to whom profiling refers to "any police-oriented action that relies on the race, ethnicity, or national origin and not merely on the behavior of an individual". Eidelson does not adopt this definition directly, nor does he present an explicit definition of racial profiling by himself. Instead, he tries to reveal the nature of racial profiling by examining two general police strategies for detecting and identifying criminals, namely, profiling (whether or not it is linked to race) and suspect description (Eidelson, 2015, pp.178-187). When police engage in profiling, they "use traits that they take to be associated with the commission of some kind of offense as sorting criteria to allocate investigative resources more efficiently" (Eidelson, 2015, p.178). When police employ a suspect description, they "use traits thought to be possessed by a specific suspect as criteria for allocating investigative resources more efficiently" (Eidelson, 2015, p.178). As Eidelson suggests, what really distinguishes profiles from suspect descriptions is their inferential structure. Suspect descriptions do not rely on generalizations about the relationship

between criminality and a trait. What a suspect description says is that a person with a particular trait is more likely to be the perpetrator in a *particular* case. By contrast, profiles suppose the existence of a *general* connection between the trait in question and criminality. What a profile says is that a person with a particular trait (a particular race in the case of racial profiling) is more likely to commit a certain *kind* of crime.

Incorporating the elements of Risse and Zeckhauser's definition and Eidelson's analysis, I formulate a definition of racial profiling as *police-oriented action that relies on an assumption about the relationship between race and criminality*.

4.2 The Problems Faced by Non-Consequentialist Arguments against Racial Profiling

Before discussing the consequentialist argument against profiling proposed by Eidelson, it is worth looking at some non-consequentialist arguments. As I discuss in Section 3, the accounts of the wrong of statistical discrimination based on disadvantaging persons for features they do not have, on merit, or on objectionable mental states, each face some dilemmas. Those accounts might nevertheless face fewer or no problems explaining racial profiling, since profiling is just one particular instance of statistical discrimination. If non-consequentialist approaches can offer a satisfactory explanation of when and why profiling is wrong, then the consequentialist account of profiling will not have any advantage on that score and may even prove to be worse.

The account based on merit clearly does not apply to the case of racial profiling, because profiling is not an instance of discrimination where merit should be the standard for selection. Turning to objectionable mental states, although individual police may be motived by racist attitudes in implementing racial profiling, either consciously or unconsciously, it is not necessary that all policies of profiling are adopted for contempt or prejudice towards racial minorities. It is

possible that a police officer who believes that black Americans have a higher crime rate and implements the policy of profiling also believes that the higher crime rate is just a result of the race's social condition rather than innate essence. The account based on objectionable mental states seem to fall short in explaining such cases, because there are no objectional mental states harbored by the discriminator.

The account based on disadvantaging persons for features they are do not have and are therefore not responsible for seems to be a more reasonable explanation for what is wrong with racial profiling. Even though the crime rate of black Americans as a group may be higher than that of white Americans, most black Americans do not commit crime. By using race as indicator for crime, the innocent black Americans who are profiled are disadvantaged in virtue of crimes they have never done. Still, as I have pointed out, this account applies to any case of statistical generalization, including those that we do not regard as problematic.

Eidelson considers one type of argument against profiling that is somewhat similar to the previous one: it contends that racial profiling is wrong because a person cannot alter, and therefore is not responsible for, his race.[9] It is unfair to impose disadvantages on persons for features they cannot change and for which they are not responsible. Eidelson invites us to consider why people believe that it would be fairer to burden a person based on mutable traits than immutable traits. He suggests that this is because we assume that when a policy burdens a person based on mutable traits, the cost to her of abandoning or suspending the trait is less than the cost of accepting the burden imposed by the policy (Eidelson, 2015, p.191). Therefore,

[9]This paper does not examine racial essentialism, but even for those who embrace such views, the claim that race cannot be altered is problematic. For those who deny the claim that race is immutable, the objection against profiling based on immutability of race does not work. What Eidelson shows here is that this objection cannot succeed even if we grant the assumption that race is immutable.

Eidelson contends, what would make profiling wrong, according to the logic of this argument, is not really the immutability of race, but the magnitude of burdens it imposes on people.

Overall, non-consequentialist accounts face problems explaining when and why racial profiling is morally troubling, for which consideration of a consequentialist account might be helpful.

4.3 The Harm Arguments against Racial Profiling

Eidelson develops an elaborate argument against racial profiling that explains the wrong of profiling in virtue of its consequences. He labels his argument as "the Harm Argument." The main idea is that racial profiling is unjustified because its costs outweigh its benefits. Eidelson (2015, p. 197) further distinguishes two forms of the Harm Argument. The *Narrow Harm Argument* focuses on the harm that profiling does to those who are subjected to police scrutiny. The *Broad Harm Argument*, in contrast, considers the harms of profiling to the people who experienced police scrutiny *and* to those who never encounter the police but are members of the race that is subject to profiling. Eidelson takes the Broad Harm Argument to be the more forceful.

Eidelson classifies the harms that profiling imposes on subjects of police scrutiny according to whether or not the harms are belief-dependent. Belief-independent harms include any delay or inconvenience caused by the scrutiny: these harms occur regardless of whether one believes that one has been profiled. Belief-dependent harms include feelings of stigma and humiliation and are dependent on the belief that one is subject to profiling, rather than on whether they are actually profiled. For example, a black male driver may reasonably believe he is racially profiled and experience feelings of stigma and humiliation when being searched by police on the road, even if the search is in fact done to every male driver including white men.

This is because, given the prevalence of racial profiling in the current society, it is reasonable for him to think that the additional scrutiny is based on race. As Eidelson put it, one's belief about whether she is profiled based on race "is more likely driven by her beliefs about the presence or absence of racial profiling in general" (Eidelson, 2015, p.199). In a society where profiling is a common practice of institutions, people of minority races are reasonably inclined to interpret any police search they are subject to as profiling. Consequently, belief-dependent harms can influence more people than those who are actually profiled.

More importantly, the harms of racial profiling are not limited to the harms against the people it affects directly but also extends to the harm of members of the target group who have never been subject to police scrutiny. This is what Eidelson highlights in the Broad Harm Argument. According to him, there are three concerns on which the Broad Harm Argument focuses (Eidelson, 2015, pp.207-208). First, a policy of profiling made known to the public conveys a demeaning and stigmatizing message about the target group, which directly injures members of the group in general, even if they are not actually profiled or believe they are profiled. Second, such a policy will influence other people's attitudes towards the target group, and those attitudes will disadvantage members of the group. Third, the harm of profiling to the target group will cause increasing resentment and hostility of members of the group toward the rest of the society, leading to racial hatred and social unrest. Eidelson believes that "the Broad Harm Argument likely furnishes a decisive objection against racial profiling in very many of the cases where it might be practiced" (Eidelson, 2015, p. 207). I focus my analysis on this argument.

The key to Eidelson's Broad Harm argument is the *social meaning* of racial profiling. The Broad Harm Argument relies on the premise that the social meaning of profiling is such that

it causally contributes to reinforcing or increasing racism. One might ask why the social meaning of profiling has such effects. Eidelson makes several points that address the question. First, the inferential structure of profiling determines its social meaning. This is shown by the contrast between profile and suspect description. Racial profiling rests on the assumption that a person with a particular race is more likely to commit a kind of crime, which is absent in suspect description (Eidelson, 2015, p.209). However, the inferential structure alone does not make the social meaning of profiling racist. After all, given that the correlation between race and crime rate does exist, which is the assumption of my discussion, what a policy of profiling "says" would apparently be just describing the world accurately. How does accurately describing the world contribute to racism?

Eidelson's explanation is that people tend to interpret the meaning of racial profiling in a racist way due to cognitive confusion. As he puts it:

> For a profile to be useful, the possession of a particular trait must only raise the epistemic probability that a person is a criminal. But the distinction between a trait's making one (epistemically) more likely to be a drug courier, and its actually making one more likely to be a drug courier, is predictably lost on many people (Eidelson, 2015, p.210).

In other words, the original assumption on which a policy of profiling is based may be different from the message that people receive from the policy. A policy of profiling can function without making any assumption about the causal relationship between race and criminality. All it depends on is the statistically positive correlation between the two variables. This point can be illustrated with Eidelson's model (Eidelson, 2015, pp.212-213). Assume that a person, x, will commit a crime at time t when $C_{t,x} > T_x$, where $C_{t,x}$ refers to the crime-conduciveness of the

person's circumstances at time t and T_x refers to the person's overall threshold for how conducive a circumstance must be for him to commit the crime. What profiling supposes is that black Americans are more likely to be in a state where $C_{t,x} > T_x$ compared with white Americans. It is neutral about whether black Americans have a higher T_x than white Americans. That is, it does not suppose that the higher crime rates of black Americans are caused by deep-seated differences internal to race. Nevertheless, human cognitive machinery is such that people are disposed to interpret races as "natural kinds" (Cosmides et al., 2003, pp. 173-179.). The psychological hypothesis that racial categories are biologically robust and homogeneous will push people towards explaining the different crime rates in terms of a different T_x between black Americans and white Americans, rather than different social circumstances they experience. Consequently, people tend to interpret the meaning of profiling as confirming the different racial essence of blacks and whites: black people are then taken to be intrinsically more prone to commit crimes. This is how a policy of racial profiling obtains a racist social meaning and causally contributes to the existence of racism in society.

4.4 Objections and Replies

In this part, I discuss some objections to Eidelson's arguments that he did not analyze but pose a challenge to his account. I argue that these objections can be replied within the framework of the account.

4.4.1 Underestimation of Benefits

The first line of objection questions whether the harm of racial profiling actually outweighs its benefit. Eidelson seems to take it for granted that once we recognize the social meaning of profiling and how it influences members of the profiled group and the rest of society, it would be self-evident that its harm is more significant than its usefulness. He writes that

"racial profiling is objectionable for the simple reason that it encourages racism, and that this harm will often outweigh whatever benefits it achieves" (Eidelson, 2015, p.214), but he does not elaborate on why it is the case. Opponents of his argument may think that Eidelson underestimates the benefits that profiling can bring. For example, Neven Sesardić argues that the price of not implementing racial profiling in the fight by the police against violent crime is very high. The situations Sesardić imagines are crimes of great violence whose prevention is matter of urgency, such as the scenario where a criminal with a deadly weapon is about to initiate an attack on a crowd, and the police cannot afford to spend the time searching every suspect. Sesardić suggests that the police officers or innocent bystanders in such a scenario are in grave danger and "their life will be put at much higher risk if the suspected attacker's race is not taken into account when deciding how to respond"[10] (Sesardić, 2018, p.998).

I think this objection is not cogent. Eidelson does not say that all forms of racial profiling are unjustifiable. His argument against profiling is a consequentialist one. Whether a particular case of profiling should be allowed is determined by considering the specific conditions. Various instances of racial profiling differ in how immediate the supposed criminal threat is, how serious the crime is, and the population it influences. Exceptional cases in extreme situations when a serious threat is posed to people's lives, such as the search for a serial killer or terrorist who might strike imminently, might still justify the use of racial profiling. Even so, Eidelson's criticism could still apply to most cases of profiling. For instance, the benefits of traffic stops and searches targeted at black people for drugs or illegal weapons can hardly outweigh their

[10] One might think this claim counterintuitive. Presumably if one is in a dangerous situation focusing on what the particular individual is doing seems more important than noting the race of the person. If the guy is waving a gun around, it is more important you notice that than what race the fellow is. To clarify, what Sesardić insists on here is not that police should not take into account factors other than race, nor that race is more important than particular features indicating the chance of violence. What he maintains is that given the same particular features, race should still be considered. For example, when two men are both waving a gun, Sesardić believes it is rational to predict the black man is more dangerous than the white man.

contribution to racism, considering that there are alternative strategies to effectively curb illegal drug or weapon possession, and the crimes do not pose an immediate threat to the public. The preventive screening of persons of a particular race at airports to catch terrorists does not seem to be beneficial enough to outweigh its harm either. Airports can strengthen security by adding security forces, installing weapon detectors, and implementing comprehensive screenings for every passenger.

4.4.2 The Meaning of Social Profiling Can Be Non-Racist

Another objection may come from those who doubt whether the social meaning of racial profiling has to be racist, as Eidelson suggests. As I have noted, the assumption of racial profiling is neutral about what causes different crime rates between races (assuming, for the sake of argument, that such differences exist). If the original meaning (contrasted with the social meaning) of a policy is the assumption on which it is based, then the original meaning of profiling is just that t*he crime rate of black Americans is higher than that of white Americans*. If the difference between the crime rates of black Americans and white Americans does exist, then the original meaning of profiling represents an accurate description of the world, which is not racist in nature.[11] If what makes racial profiling objectionable is not its original meaning, but people's interpretation of it, then instead of banning the policy itself, the government should make efforts to correct people's understanding of it.

Nevertheless, this policy suggestion would not be practical. As Eidelson (2015, p.212) points out and is also supported by psychological research (Quattrone & Jones,1980; Judd & Park, 1991; McGarty, 1999), there is an underlying cognitive mechanism that naturally pushes

[11] Here I suppose the meaning of crime is not biased by racism. In a racist society, it is possible for the definition of what counts as crime to be skewed in a way that unjustly disadvantages racial minorities.

people to think of a racial outgroup as homogenous. This mechanism determines that white Americans are disposed to view black Americans as similar in their tendency to commit crimes, rather than regard the higher crime rates as reflecting mere differences in people's circumstances. It would be difficult to reverse the human psychological machinery that is likely to be carved by genes. Besides the cognitive mechanism, the social context is also important for determining the social meaning of racial profiling. Consider gender profiling targeted at men based on the assumption that men are more likely to commit crimes than women. Since gender is also often regarded as a robust category, people also tend to interpret the meaning of gender profiling as suggesting that men are intrinsically more violent and crime-prone than women. Still, few people find gender profiling as objectionable as racial profiling. This is because men as a group are regarded as dominant and superior. The crime-proneness of men is not linked to inferiority. Instead, it may be seen as an expression of suitable aggressiveness and power, a typical part of the conventional image of what men ought to be. By contrast, there is widespread disrespect towards black Americans, who are perceived to be inferior to white Americans. In such context, violence is regarded as confirmation of black Americans' inferiority; thus, "affirming the notion that race is predictive of criminality…is *conventionally* disrespectful," even in the absence of disrespectful intention (Eidelson, 2015, p.186, Emphasis in original). Therefore, to correct the social meaning of racial profiling, we would also need to change the social context and reduce the existing discrimination against black Americans. This would be a long-term and complex program.[12]

[12] In fact, a strong case can be made that by the time that discrimination against black people is eliminated, there will not likely to be a great difference between the crime rates of blacks and whites. Then there would be no need to correct the social meaning of racial profiling, because there would be no need to implement racial profiling at all.

4.4.3 Racial profiling and Perception

A third kind of objection, one that I think most forceful, is based on the distinction between the practice of racial profiling and the perception of it. As Eidelson points out, the major harm of profiling is belief-dependent. Without the belief that one's own group is being the target of profiling, members of the profiled groups are not likely to have the feeling of hurt, stigma and resentment. Without the belief that certain minority groups are the target of profiling, members of the other groups are not likely to reinforce their racist attitudes against the minority groups. That is to say, most of the harmful consequences related with profiling depend on people's perception of profiling, rather than profiling itself. If what matters is whether people believe profiling is employed, then is not it possible to preserve profiling itself without producing high social costs? In fact, one can doubt whether Eidelson's argument is a case against racial profiling at all. Most of the harms that he describes (except the delay and inconvenience caused by additional scrutiny) would only occur if there is public perception of profiling. In other words, whether the Harm Argument applies is independent of whether there actually is profiling at all. So it seems that what Eidelson shows is not that racial profiling itself is wrong, but that the public perception of profiling causes wrongful harm.

I agree that Eidelson's argument is, strictly speaking, against the perception rather than the practice of profiling, but it does not mean his conclusion about the immorality of profiling in the *current society* fails. Given the social conditions of the current US, most racial profiling cannot be morally justified and will remain unjustifiable at least in the near term. These social conditions include several aspects. First, the history of the US is such that black Americans have suffered injustice and oppression economically, socially and institutionally. Second, there are prevalent discriminatory and disrespectful attitudes towards black Americans that conceive them

as violent, dangerous and inferior. Despite efforts of civil right activists to alter these attitudes, the attitudes remain entrenched and will continue to exist in the long run. Third, although it is possible to separate the practice of racial profiling and the perception of it in theory, in the real world, perception and practice are inextricably intertwined. People invariably notice that one race is being searched by police more than the other race. If government were to falsely deny that it practices racial profiling, the denial would be met by widespread incredulity and criticism. In sum, even though the harmful consequences of profiling are largely mediated by social perceptions, entrenched social realities mean that the elimination of such consequences cannot be reasonably expected anytime soon.

5 CONCLUSION

Overall, my thesis argues that a context-based consequentialist account is a promising explanation for the wrong of statistical discrimination. The distinct-wrong approach to understanding the wrong of statistical discrimination, holding that it consists in the use of statistical evidence, is misguided because it cannot demarcate morally wrongful statistical discrimination from the justifiable use of statistical generalizations. Among the different versions of the general-wrong approach, I have maintained that a consequentialist account is preferable to the accounts based on merit or objectionable mental states. Specifically, I argued that the wrong of statistical discrimination should be explained in terms of its effect of contributing to and reinforcing the unjust subordination of disadvantaged social groups in a given social context.

To provide further support for my account, I showed that the context-based consequentialist account can provide an illuminating way to reveal the morally troubling features

of racial profiling by defending Eidelson's argument against profiling, which is a version of consequentialist account. The wrongfulness of racial profiling is determined by the harmful consequences that it produces, in particular, the perpetuation and reinforcement of racism.

As a final remark, I would like to make some clarifications on my account. First, the account relies on some controversial empirical claims that not all reasonable people will agree with, such as the assumptions about the impact of racial profiling on society. My argument would be weakened if those empirical assumptions prove to be false. Still, the assumptions are plausible and reflect widely held understandings of US society among informed people. Second, what I offer is only a preliminary defense for the context-based consequentialist account. My argument does not rebut alternative accounts of the wrong of discrimination or establish that my account is the only plausible one. My goal is just to show that a context-based consequentialist account has notable strengths compared to alternatives. Third, my argument does not preclude the possibility that there is more than one source of the moral wrong of wrongful statistical discrimination. It is possible that some instances of statistical discrimination are wrong for both consequentialist and deontological reasons. My claim here is that the consequentialist account gives a reasonably convincing explanation of the wrongfulness of at least some notable instances of wrongful statistical discrimination.

BIBLIOGRAPHY

Alexander, L. (1992). What makes wrongful discrimination wrong? Biases, preferences, stereotypes, and proxies. *University of Pennsylvania Law Review*, *141*(1), 149-219.

Altman, A. (2020). *Discrimination*. Stanford Encyclopedia of Philosophy. Retrieved 19 April 2020, from https://plato.stanford.edu/entries/discrimination/.

Anderson, E. S., & Pildes, R. H. (2000). Expressive theories of law: A general restatement. *University of Pennsylvania Law Review*, *148*(5), 1503-1575.

Arneson, R. J. (2006). What is wrongful discrimination. *San Diego Law Review*, 43, 775-805.

Brubaker, R., Loveman, M., & Stamatov, P. (2004). Ethnicity as cognition. *Theory and Society*, *33*(1), 31-64.

Cavanagh, M. (2002). *Against equality of opportunity*. Clarendon Press.

Colyvan, M., Regan, H. M., & Ferson, S. (2001). Is it a crime to belong to a reference class. *Journal of Political Philosophy*, *9*(2), 168-181.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*(4), 173-179.

Eidelson, B. (2015). *Discrimination and disrespect*. Oxford University Press.

Glasgow, J. (2009). Racism as disrespect. *Ethics*, *120*(1), 64-93.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.

Hayes, P., Poel, I., & Steen, M. (2020). Algorithms and values in justice and security. *AI & Society*, *35*(3),1-23.

Hellman, D. (2018) Discrimination and social meaning. In Lippert-Rasmussen (ed.), *The Routledge handbook of the ethics of discrimination.* (pp. 97–118). Routledge.

Hellman, D. (2008). *When is discrimination wrong?*. Harvard University Press.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology*, *61*(3), 366-379.

Kim, P. T. (2016). Data-driven discrimination at work. *William. & Mary Law Review.*, 58(3), 857-936.

Lever, A. (2005). Why racial profiling is hard to justify: A response to Risse and Zeckhauser. *Philosophy & Public Affairs, 33*(1), 94-110.

Lippert-Rasmussen, K. (2006). Racial profiling versus community. *Journal of Applied Philosophy*, *23*(2), 191-205.

Lippert-Rasmussen, K. (2014). *Born free and equal?: A philosophical inquiry into the nature of discrimination*. Oxford University Press.

Lippert-Rasmussen, K. (2011). "We are all different": Statistical discrimination and the right to be treated as an individual. *The Journal of Ethics*, *15*(1-2), 47-59.

Lippert-Rasmussen, K. (2007). Nothing personal: On statistical discrimination. *Journal of Political Philosophy*, *15*(4), 385-403.

Risse, M., & Zeckhauser, R. (2004). Racial profiling. *Philosophy & Public Affairs*, *32*(2), 131-170.

McGarty, C. (1999). *Categorization in social psychology*. Sage.

MacKinnon, C. A. (1987). *Feminism unmodified: Discourses on life and law*. Harvard University Press.

Miller, D. (2001). *Principles of social justice*. Harvard University Press.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659-661

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social psychology*, *38*(1), 141-152.

Rodgers, W. M. (2009). *Handbook on the economics of discrimination*. Edward Elgar Publishing.

Scanlon, T. M. (2009). *Moral dimensions*. Harvard University Press.

Scanlon, T. (1998). *What we owe to each other*. Harvard University Press.

Sesardić, N. (2018). Is racial profiling a legitimate strategy in the fight against violent crime?.*Philosophia*, *46*(4), 981-999.

Theunissen, G., Verbruggen, M., Forrier, A., & Sels, L. (2011). Career sidestep, wage setback? The impact of different types of employment interruptions on wages. *Gender, Work & Organization*, *18*, 110-131.

Thomsen, F. K. (2013). But some groups are more equal than others: a critical review of the group criterion in the concept of discrimination. *Social Theory and Practice*, *39*(1), 120-146.